SUCCESS STORY

scch { }

Software Competence Center Hagenberg GmbH

Programme: COMET - Competence Centers for Excellent Technologies

Programme line: COMET-Zentrum (K1)

Duration: 2023-2026

THE SEMANTIC DATA QUALITY WORKFLOW

From Rigid Metrics to Smart KPIs

—

Together with Robert Bosch AG, we developed a Semantic Data Quality workflow that enables measurement of data quality based on domain knowledge which is represented as a knowledge graph.

Modern manufacturing ecosystems and supply chains are embedded in smart infrastructures, forming a vast, interconnected data and information network. Knowledge Graphs (KGs) have emerged as a powerful technology to manage this complexity using ontologies to provide rich, domain-specific context to raw data.

However, this data-centered approach faces a critical vulnerability: data quality. The principle of "garbage in, garbage out" has never been more relevant. If the data is incomplete, inconsistent, or inaccurate, the decisions based on it will be flawed, leading to inefficient production, compliance failures, and costly errors.

The challenge lies in the fact that data quality cannot be addressed through a universal approach; it is inherently dependent on context. Sensor readings that are updated every ten seconds may be considered “complete” for some use cases (e.g. non-critical monitoring), but “incomplete” for others (e.g., real-time safety control).

Previous attempts to solve this challenge used ontologies [1], such as the Data Quality Definition Ontology (DQD), to define data quality metrics. While a good first step, this approach had a fundamental flaw: it forced developers to hard-code the context (the "how" to check) directly into the metric.

Defining a completeness check for two similar sensor streams, such as the completeness of a temperature

sensor and a pressure sensor, would result in two entirely new, almost identical, metric definitions. This approach is rigid, time-consuming, and severely limits the reusability of metrics, creating a bottleneck for scaling data quality assurance.

Instead of such rigid definitions, DQD 2.0 introduces a novel method for *dynamic, knowledge-graph-based parameters*. A data quality metric is defined only once, the context (the "how" to check) is passed to this metric as a flexible parameter retrieved from the knowledge graph at runtime [2, 3]. Our patented Data Quality workflow using DQD 2.0 performs three key actions:

1. **Defines:** It uses a data quality ontology to define data quality metrics in a flexible and reusable way.

2. **Executes:** It evaluates these dynamic data quality metrics against raw data entities, represented as tabular data or also as knowledge graphs.

3. **Annotates:** It stores data quality results in a knowledge graph, linking quality metrics and their results directly to the entities that were checked

Impact and Effects

Massive Reusability and Scalability. A quality assurance team can apply a single defined data quality metric to an arbitrary number of different contexts dynamically. This eliminates the need to manually create and maintain a huge number of redundant metric definitions, saving time and reducing the potential for human error.

> **The core innovation of our workflow is its ability to decouple the data quality metric and the domain knowledge required to evaluate the metric.**

Data Quality becomes Accessible. Another significant impact is the annotation of results back onto the graph (see Fig. 1). Previously, data quality results were often siloed in separate reports or databases, disconnected from the data itself. The data quality results are directly connected with the data, making them easy to find, analyze, and act upon. This allows for the creation of real-time dashboards, automated alerts, and a continuous improvement loop for data governance.

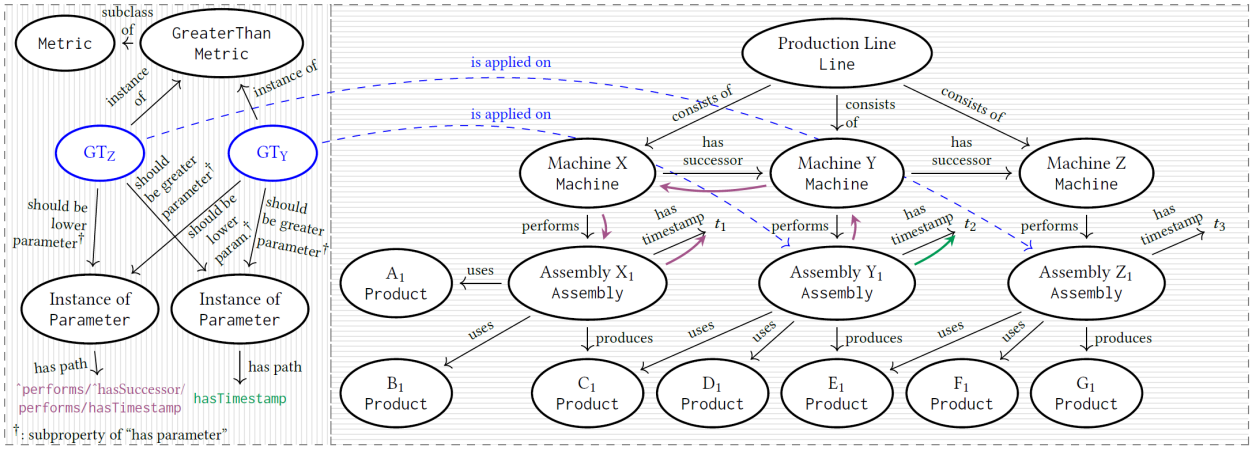

Figure 1: Example graph resulting from applying our KG based Data Quality Workflow

Related Publications

[1] J. Schrott, R. Meindl, C. Lettner, W. Wöß, and L. Ehrlinger, DQD: The Data Quality Definition Ontology, in *Metadata and Semantic Research*, vol. 2048, E. Garoufallou and F. Sartori, Eds., in Communications in Computer and Information Science, vol. 2048. , Cham: Springer Nature Switzerland, 2024, pp. 291–297. doi: 10.1007/978-3-031-65990-4_27.

[2] J. Schrott, R. Meindl, C. Lettner, S. Hammer, and M. Leitner, Dynamic Knowledge Graph-based Measurement of Data Quality, in *Proceedings of the VLDB Endowment. ISSN*, [Online]. Available: https://www.vldb.org/2025/Workshops/VLDB-Workshops-2025/QDB/QDB25_3.pdf

[3] Patent Application DE102024129324A1: *Method and device for diagnosing or monitoring the (data) quality of a production plant, its production units and products*; https://patents.google.com/patent/DE102024129324A1/en

Project Partner

- Software Competence Center Hagenberg
- Robert Bosch AG

Contact

Mag. Christian Lettner
Senior Research Project Manager Data Science
+43 50 343 837
christian.lettner@scch.at

This success story was provided by the Software Competence Center Hagenberg GmbH and by the mentioned project partners for the purpose of being published. Software Competence Center Hagenberg is a COMET Centre within the COMET – Competence Centers for Excellent Technologies Programme and funded by BMIMI, BMWET, and the Province of Upper Austria. The COMET Programme is managed by FFG. Further information on COMET: www.ffg.at/comet

Federal Ministry
Innovation, Mobility
and Infrastructure
Republic of Austria

Federal Ministry
Economy, Energy
and Tourism
Republic of Austria

Austrian Research Promotion Agency
Sensengasse 1, A-1090 Vienna
P +43 (0) 5 77 55 - 0
office@ffg.at
www.ffg.at